

# A Distinguisher-Based Attack of a Homomorphic Encryption Scheme Relying on Reed-Solomon Codes

Valérie Gauthier<sup>1</sup>, Ayoub Otmani<sup>1</sup> and Jean-Pierre Tillich<sup>2</sup>

<sup>1</sup> GREYC - Université de Caen - Ensicaen

Boulevard Maréchal Juin, 14050 Caen Cedex, France.

valerie.gauthier01@unicaen.fr, ayoub.otmani@unicaen.fr,

<sup>2</sup> SECRET Project - INRIA Rocquencourt

Domaine de Voluceau, B.P. 105 78153 Le Chesnay Cedex - France

jean-pierre.tillich@inria.fr

**Abstract.** Bogdanov and Lee suggested a homomorphic public-key encryption scheme based on error correcting codes. The underlying public code is a modified Reed-Solomon code obtained from inserting a zero submatrix in the Vandermonde generating matrix defining it. The columns that define this submatrix are kept secret and form a set  $L$ . We give here a distinguisher that detects if one or several columns belong to  $L$  or not. This distinguisher is obtained by considering the code generated by component-wise products of codewords of the public code (the so called “square code”). This operation is applied to punctured versions of this square code obtained by picking a subset  $I$  of the whole set of columns. It turns out that the dimension of the punctured square code is directly related to the cardinality of the intersection of  $I$  with  $L$ . This allows an attack which recovers the full set  $L$  and which can then decrypt any ciphertext.

## 1 Introduction

The concept of homomorphic encryption was first proposed in 1978 in [RAD78]. But it took more than three decades to see the first scheme of this kind [Gen09]. It is based on ideal lattices. Since then several proposals have been made, most of them rely on lattice theory. One challenging issue is to come up with a homomorphic encryption scheme using different security assumptions. Recently, the first symmetric homomorphic encryption scheme based on error-correcting codes was proposed in [AAPS11]. This work was then followed by [BL12] which can be considered as the first public-key homomorphic scheme based on coding theory. This particular cryptosystem heavily relies on properties of Reed-Solomon codes. These codes have been suggested for the first time in a public-key cryptosystem in [Nie86] but it was shown to be insecure in [SS92]. The attack recovers the underlying Reed-Solomon allowing the decoding of any encrypted data obtained from a McEliece-type cryptosystem based on them. The McEliece cryptosystem [McE78] on the other hand uses Goppa codes. Since its apparition, it has withstood many attacks and after more than thirty years now, it still belongs to the very few unbroken public key cryptosystems. This situation substantiates the claim that inverting the encryption function, and in particular recovering the private key from public data, is intractable.

No significant breakthrough has been observed with respect to the problem of recovering the private key [Gib91,LS01]. This has led to claim that the generator matrix of a binary Goppa code does not disclose any visible structure that an attacker could exploit. This is strengthened by the fact that Goppa codes share many characteristics with random codes: for instance they asymptotically meet the Gilbert-Varshamov bound, they typically have a trivial permutation group, *etc.* Hence, the hardness of the Goppa code distinguishing problem, which asks whether a Goppa code can be distinguished from a random code, has become a classical belief in code-based cryptography, and

as a consequence, a mandatory assumption to prove semantic security in the random oracle model [NIKM08], CCA2 security in the standard model [DMQN09] and security in the random oracle model against existential forgery [CFS01,Dal07] of the signature scheme [CFS01].

In [FGO<sup>+</sup>11], an algorithm that manages to distinguish between a random code and a Goppa code has been introduced. This work without undermining the security of [McE78] prompts to wonder whether it would be possible to devise an attack based on such a distinguisher. It was found out in [MCP12] that our distinguisher [FGO<sup>+</sup>11] has an equivalent but simpler description in terms of the component-wise product of codes. This notion was first put forward in coding theory to unify many different algebraic decoding algorithms [Pel92,Köt92]. This distinguisher is even more powerful in the case of Reed-Solomon codes than for Goppa codes because, whereas for Goppa codes it is only successful for rates close to 1, it can distinguish Reed-Solomon codes of any rate from random codes. In the specific case of [BL12], the underlying public code is a modified Reed-Solomon code obtained from inserting a zero submatrix in the Vandermonde generating matrix defining it and in this case our distinguisher leads to an attack. We present namely in this paper a key-recovery attack on the Bogdanov-Lee homomorphic scheme based on the version of our distinguisher presented in [MCP12]. Our attack runs in polynomial time and is efficient: it only amounts to calculate the ranks of certain matrices derived from the public key.

More precisely, in their cryptosystem the columns that define the zero submatrix are kept secret and form a set  $L$ . We give here a distinguisher that detects if one or several columns belong to  $L$  or not. This distinguisher is obtained by considering the code generated by component-wise products of codewords of the public code (the so called “square code”). This operation is applied to punctured versions of this square code obtained by picking a subset  $I$  of the whole set of columns. It turns out that the dimension of the punctured square code is directly related to the cardinality of the intersection of  $I$  with  $L$ . This allows an attack which recovers the full set  $L$  and which can then decrypt any ciphertext.

It should also be pointed out that the properties of Reed-Solomon codes with respect to the component-wise product of codes have already been used to cryptanalyze a McEliece scheme based on subcodes of Reed-Solomon codes [Wie10]. The use of this product is nevertheless different in [Wie10] from the way we use it here. Note also that our attack is not an adaptation of the Sidelnikov and Shestakov approach [SS92]. Our approach is completely new: it illustrates how a distinguisher that detects an abnormal behaviour can be used to recover the private key.

In Section 2 we recall important notions from coding theory. In Section 3 we introduce the cryptosystem and in Section 4 we present the key recovery attack.

## 2 Reed-Solomon codes and the square code

We recall in this section a few relevant results and definitions from coding theory and bring in the fundamental notion which is used in the attack, namely the square code. A linear *code*  $\mathcal{C}$  of *length*  $n$  and *dimension*  $k$  over a finite field  $GF(q)$  of  $q$  elements is a subspace of dimension  $k$  of the full space  $GF(q)^n$ . It is generally specified by a full-rank matrix called a generator matrix which is a  $k \times n$  matrix  $\mathbf{G}$  (with  $k \leq n$ ) over  $GF(q)$  whose rows span the code:

$$\mathcal{C} = \left\{ \mathbf{u}\mathbf{G} \mid \mathbf{u} \in GF(q)^k \right\}.$$

It can also be specified by a *parity-check* matrix  $\mathbf{H}$ , which is a matrix whose right kernel is equal to the code, that is

$$\mathcal{C} = \left\{ \mathbf{x} \in GF(q)^n \mid \mathbf{H}\mathbf{x}^T = 0 \right\},$$

where  $\mathbf{x}^T$  stands for the column vector which is the transpose of the row vector  $\mathbf{x}$ . The *rate* of the code is given by the ratio  $\frac{k}{n}$ . Code-based public-key cryptography focuses on linear codes that have a polynomial time decoding algorithm. The role of decoding algorithms is to correct errors of prescribed weight. We say that a decoding algorithm corrects  $t$  errors if it recovers  $\mathbf{u}$  from the knowledge of  $\mathbf{u}\mathbf{G} + \mathbf{e}$  for all possible  $\mathbf{e} \in \mathbb{F}_q^n$  of weight at most  $t$ .

Reed-Solomon codes form a special case of codes with a very powerful low complexity decoding algorithm. It will be convenient to use the definition of Reed-Solomon codes and generalized Reed-Solomon codes as *evaluation codes*

**Definition 1 (Reed-Solomon code and generalized Reed-Solomon code).** *Let  $k$  and  $n$  be integers such that  $1 \leq k < n \leq q$  where  $q$  is a power of a prime number. Let  $\mathbf{x} = (x_1, \dots, x_n)$  be an  $n$ -tuple of distinct elements of  $GF(q)$ . The Reed-Solomon code  $\mathbf{RS}_k(\mathbf{x})$  of dimension  $k$  is the set of  $(p(x_1), \dots, p(x_n))$  when  $p$  ranges over all polynomials of degree  $\leq k - 1$  with coefficients in  $GF(q)$ . The generalized Reed-Solomon code  $\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})$  of dimension  $k$  is associated to a couple  $(\mathbf{x}, \mathbf{y}) \in GF(q)^n \times GF(q)^n$  where  $\mathbf{x}$  is chosen as above and the entries  $y_i$  are arbitrary non zero elements in  $GF(q)$ . It is defined as the set of  $(y_1p(x_1), \dots, y_np(x_n))$  where  $p$  ranges over all polynomials of degree  $\leq k - 1$  with coefficients in  $GF(q)$ .*

Generalized Reed-Solomon codes are quite important in coding theory due to the conjunction of several factors such as :

- (i) their minimum distance  $d$  is maximal among all codes of the same dimension and length since they are MDS codes (their distance is equal to  $n - k + 1$ ),
- (ii) they can be efficiently decoded in polynomial time when the number of errors is less than or equal to  $\lfloor \frac{d-1}{2} \rfloor = \lfloor \frac{n-k}{2} \rfloor$ .

It has been suggested to use them in a public-key cryptosystem for the first time in [Nie86] but it was discovered that this scheme is insecure in [SS92]. Sidelnikov and Shestakov namely showed that it is possible to recover in polynomial time for any generalized Reed-Solomon code a possible couple  $(\mathbf{x}, \mathbf{y})$  which defines it. This is all what is needed to decode efficiently such codes and is therefore enough to break the Niederreiter cryptosystem suggested in [Nie86] or a McEliece type cryptosystem [McE78] when Reed-Solomon are used instead of Goppa codes.

We could not find a way to adapt the Sidelnikov and Shestakov approach for cryptanalyzing the Bogadnov and Lee cryptosystem. However a Reed-Solomon displays a quite peculiar property with respect to the component-wise product which is denoted by  $\mathbf{a} \star \mathbf{b}$  for two vectors  $\mathbf{a} = (a_1, \dots, a_n)$  and  $\mathbf{b} = (b_1, \dots, b_n)$  and which is defined by  $\mathbf{a} \star \mathbf{b} \stackrel{\text{def}}{=} (a_1b_1, \dots, a_nb_n)$ . This can be seen by bringing in the following definition

**Definition 2 (star product of two codes, square code).** *Let  $\mathcal{A}$  and  $\mathcal{B}$  be two codes of length  $n$ . The star product code denoted by  $\langle \mathcal{A} \star \mathcal{B} \rangle$  of  $\mathcal{A}$  and  $\mathcal{B}$  is the vector space spanned by all products  $\mathbf{a} \star \mathbf{b}$  where  $\mathbf{a}$  and  $\mathbf{b}$  range over  $\mathcal{A}$  and  $\mathcal{B}$  respectively. When  $\mathcal{B} = \mathcal{A}$ ,  $\langle \mathcal{A} \star \mathcal{A} \rangle$  is called the square code of  $\mathcal{A}$  and is denoted by  $\langle \mathcal{A}^2 \rangle$ .*

It is clear that  $\langle \mathcal{A} \star \mathcal{B} \rangle$  is also generated by the  $\mathbf{a}_i \star \mathbf{b}_j$ 's where the  $\mathbf{a}_i$ 's and the  $\mathbf{b}_j$ 's form a basis of  $\mathcal{A}$  and  $\mathcal{B}$  respectively. Therefore

**Proposition 1.**

$$\dim(\langle \mathcal{A} \star \mathcal{B} \rangle) \leq \dim(\mathcal{A})\dim(\mathcal{B}).$$

We expect that the square code when applied to a random linear code should be a code of dimension of order  $\min \left\{ \binom{k+1}{2}, n \right\}$ . Actually by using the proof technique of [FGO<sup>+</sup>11] it can be shown for instance that with probability going to 1 as  $k$  tends to infinity, the square code is of dimension

$\min \left\{ \binom{k+1}{2}(1 + o(1)), n \right\}$  when  $k$  is of the form  $k = o(n^{1/2})$ . On the other hand generalized Reed Solomon codes behave in a completely different way

**Proposition 2.**  $\langle \mathbf{GRS}_k(\mathbf{x}, \mathbf{y})^2 \rangle = \mathbf{GRS}_{2k-1}(\mathbf{x}, \mathbf{y} \star \mathbf{y})$ .

This follows immediately from the definition of a generalized Reed Solomon code as an evaluation code since the star product of two elements  $\mathbf{c} = (y_1p(x_1), \dots, y_np(x_n))$  and  $\mathbf{c}' = (y_1q(x_1), \dots, y_nq(x_n))$  of  $\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})$  where  $p$  and  $q$  are two polynomials of degree at most  $k-1$  is of the form

$$\mathbf{c} \star \mathbf{c}' = (y_1^2p(x_1)q(x_1), \dots, y_n^2p(x_n)q(x_n)) = (y_1^2r(x_1), \dots, y_n^2r(x_n))$$

where  $r$  is a polynomial of degree  $\leq 2k-2$ . Conversely, any element of the form  $(y_1^2r(x_1), \dots, y_n^2r(x_n))$  where  $r$  is a polynomial of degree less than or equal to  $2k-1$  is a linear combination of star products of two elements of  $\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})$ .

This proposition shows that the square code is only of dimension  $2k-1$  when  $2k-1 \leq n$ , which is quite unusual. This property can also be used in the case  $2k-1 > n$ . To see this, consider the dual of the Reed-Solomon code. The *dual*  $\mathcal{C}^\perp$  of a code  $\mathcal{C}$  of length  $n$  over  $GF(q)$  is defined by

$$\mathcal{C}^\perp = \{ \mathbf{x} \in GF(q)^n \mid (\mathbf{x}, \mathbf{y}) = 0, \mathbf{y} \in \mathcal{C} \},$$

where  $(\mathbf{x}, \mathbf{y}) = \sum x_i y_i$  stands for the standard inner product between elements of  $GF(q)^n$ . The dual of a generalized Reed-Solomon code is itself a generalized Reed-Solomon code, see [MS86, Theorem 4, p.304]

**Proposition 3.**

$$\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})^\perp = \mathbf{GRS}_{n-k}(\mathbf{x}, \mathbf{y}')$$

where the length of  $\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})$  is  $n$  and  $\mathbf{y}'$  is a certain element of  $GF(q)^n$  depending on  $\mathbf{x}$  and  $\mathbf{y}$ .

Therefore when  $2k-1 > n$  a Reed-Solomon code  $\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})$  can also be distinguished from a random linear code of the same dimension by computing the dimension of  $\langle (\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})^\perp)^2 \rangle$ . We have in this case

$$\langle (\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})^\perp)^2 \rangle = \langle \mathbf{GRS}_{n-k}(\mathbf{x}, \mathbf{y}')^2 \rangle = \langle \mathbf{GRS}_{2n-2k-1}(\mathbf{x}, \mathbf{y}' \star \mathbf{y}') \rangle$$

and we obtain a code of dimension  $2n - 2k - 1$ .

The star product of two codes is the fundamental notion used in the decoding algorithm based on an error correcting pair [Pel92, Köt92] which unifies common ideas to many algebraic decoding algorithms. It has been used for the first time to cryptanalyze a McEliece scheme based on subcodes of Reed-Solomon codes [Wie10]. The use of the star product is nevertheless different in [Wie10] from the way we use it here. In this paper, the star product is used to identify for a certain subcode  $\mathcal{C}$  of a generalized Reed-Solomon code  $\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})$  a possible pair  $(\mathbf{x}, \mathbf{y})$ . This is achieved by computing  $\langle \mathcal{C}^2 \rangle$  which in the case which is considered turns out to be equal to  $\langle \mathbf{GRS}_k(\mathbf{x}, \mathbf{y})^2 \rangle$  which is equal to  $\mathbf{GRS}_{2k-1}(\mathbf{x}, \mathbf{y} \star \mathbf{y})$ . The Sidelnikov and Shestakov is then used on  $\langle \mathcal{C}^2 \rangle$  to recover a possible  $(\mathbf{x}, \mathbf{y} \star \mathbf{y})$  pair to describe  $\langle \mathcal{C}^2 \rangle$  as a generalized Reed-Solomon code. From this, a possible  $(\mathbf{x}, \mathbf{y})$  pair for which  $\mathcal{C} \subset \mathbf{GRS}_k(\mathbf{x}, \mathbf{y})$  is deduced.

### 3 The Bogdanov-Lee Cryptosystem

The cryptosystem proposed by Bogdanov and Lee in [BL12] is a public-key homomorphic encryption scheme based on linear codes. It encrypts a plaintext  $m$  from  $GF(q)$  into a ciphertext  $\mathbf{c}$  that belongs to  $GF(q)^n$  where  $n$  is a given integer. The key generation requires a non-negative integer  $\ell$  such that  $3\ell < n$  and a subset  $L$  of  $\{1, \dots, n\}$  of cardinality  $3\ell$ . A set of  $n$  distinct elements  $x_1, \dots, x_n$  from  $GF(q)$  are generated at random. They serve to construct a  $k \times n$  matrix  $\mathbf{G}$  whose  $i$ -th column  $\mathbf{G}_i^T$  ( $1 \leq i \leq n$ ) is defined by:

$$\mathbf{G}_i^T \stackrel{\text{def}}{=} \begin{cases} (x_i, x_i^2, \dots, x_i^\ell, 0, \dots, 0) & \text{if } i \in L \\ (x_i, x_i^2, \dots, x_i^\ell, x_i^{\ell+1}, \dots, x_i^k) & \text{if } i \notin L \end{cases}$$

where the symbol  $T$  stands for the transpose.

In other words, when  $L$  is the set  $\{1, \dots, 3\ell\}$ ,  $\mathbf{G}$  is the following matrix:

$$\begin{pmatrix} x_1 & \dots & x_{3\ell} & x_{3\ell+1} & \dots & x_n \\ \vdots & & \vdots & \vdots & & \vdots \\ x_1^\ell & \dots & x_{3\ell}^\ell & x_{3\ell+1}^\ell & \dots & x_n^\ell \\ 0 & \dots & 0 & x_{3\ell+1}^{\ell+1} & \dots & x_n^{\ell+1} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & x_{3\ell+1}^k & \dots & x_n^k \end{pmatrix}.$$

The cryptosystem is now defined as follows.

- **Secret key:**  $(L, \mathbf{G})$ .
- **Public key:**  $\mathbf{P} \stackrel{\text{def}}{=} \mathbf{S}\mathbf{G}$  where  $\mathbf{S}$  is a  $k \times k$  random invertible matrix over  $GF(q)$ .
- **Encryption:** the ciphertext  $\mathbf{c} \in GF(q)^n$  of a plaintext  $m \in GF(q)$  is obtained by picking  $\mathbf{x}$  in  $GF(q)^k$  uniformly at random and  $\mathbf{e}$  in  $GF(q)^n$  by choosing its components according to a certain distribution  $\tilde{\eta}$ , then computing  $\mathbf{c} \stackrel{\text{def}}{=} \mathbf{x}\mathbf{P} + m\mathbf{1} + \mathbf{e}$  where  $\mathbf{1} \in GF(q)^n$  is the all-ones row vector.
- **Decryption:** the linear system (1) is solved for  $\mathbf{y} \stackrel{\text{def}}{=} (y_1, \dots, y_n) \in GF(q)^n$ :

$$\begin{cases} \mathbf{G}\mathbf{y}^T = 0 \\ \sum_{i \in L} y_i = 1 \\ y_i = 0 \text{ for all } i \notin L. \end{cases} \quad (1)$$

The plaintext is  $m = \sum_{i=1}^n y_i c_i$ .

The decryption algorithm will output the correct plaintext when  $\ell$  and  $n$  are chosen such that the entry  $e_i$  at position  $i$  of the error vector is zero when  $i \in L$ . The distribution  $\eta$  which is used to draw at random the coordinates of  $\mathbf{e}$  is chosen such that this property holds with very large probability. To check the correctness of the algorithm when this property on  $\mathbf{e}$  holds, notice that the linear system (1) has  $3\ell$  unknowns and  $\ell + 1$  equations and since it is by construction of rank

$\ell + 1$ , it always admits at least one solution. Then observe that

$$\begin{aligned}
\sum_{i=1}^n y_i c_i &= (\mathbf{xP} + m\mathbf{1} + \mathbf{e})\mathbf{y}^T \\
&= (\mathbf{xP} + m\mathbf{1})\mathbf{y}^T \quad (\text{since } e_i = 0 \text{ if } i \in L \text{ and } y_i = 0 \text{ if } i \notin L) \\
&= \mathbf{xSGy}^T + m \sum_{i=1}^n y_i \\
&= m \quad (\text{since } \mathbf{Gy}^T = 0 \text{ and } \sum_{i=1}^n y_i = 1).
\end{aligned}$$

The parameters  $k, q, \ell$  and the noise distribution  $\tilde{\eta}$  are chosen such as

- $q = \Omega(2^{n^\alpha})$ ;
- $k = \Theta(n^{1-\alpha/8})$ ;
- the noise distribution  $\tilde{\eta}$  is the  $q$ -ary symmetric channel with noise rate  $\eta = \Theta(1/n^{1-\alpha/4})$ , that is  $\text{Prob}(e_i = 0) = 1 - \eta$  and  $\text{Prob}(e_i = x) = \frac{\eta}{q-1}$  for any  $x$  in  $GF(q)$  different from zero;
- $\ell = \Theta(n^{\alpha/4})$ ;

where  $\alpha$  is some constant in the range  $(0, \frac{1}{4}]$ . It is readily checked that the probability that  $e_i \neq 0$  for  $i \in L$  is vanishing as  $n$  goes to infinity since it is upper-bounded by  $\eta\ell = \Theta\left(\frac{n^{\alpha/4}}{n^{1-\alpha/4}}\right) = \Theta(n^{-1+\alpha/2}) = o(1)$ .

## 4 An efficient attack on the Bogdanov-Lee homomorphic cryptosystem

### 4.1 Outline

The attack consists in first recovering the secret set  $L$  and from here finds directly a suitable vector  $\mathbf{y}$  by solving the system

$$\begin{cases} \mathbf{Py}^T = 0 \\ \sum_{i \in L} y_i = 1 \\ y_i = 0 \text{ for all } i \notin L. \end{cases} \quad (2)$$

Indeed, requiring that  $\mathbf{Py}^T = 0$  is equivalent to  $\mathbf{SGy}^T = 0$  and since  $\mathbf{S}$  is invertible this is equivalent to the equation  $\mathbf{Gy}^T = 0$ . Therefore System (2) is equivalent to the “secret” system (1). An attacker may therefore recover  $m$  without even knowing  $\mathbf{G}$  just by outputting  $\sum_i y_i c_i$  for any solution  $\mathbf{y}$  of (2). In the following subsection, we will explain how  $L$  can be recovered from  $\mathbf{P}$  in polynomial time.

### 4.2 Recovering $L$

Our attack relies heavily on the fact that the public matrix may be viewed as a the generator matrix of a code  $\mathcal{C}$  which is quite close to a generalized Reed-Solomon code (or to a Reed-Solomon if a row consisting only of 1’s is added to it). Notice that any punctured version of the code has also this property (a punctured code consists in keeping only a fixed subset of positions in a codeword). More precisely, let us introduce

**Definition 3.** For any  $I \subset \{1, \dots, n\}$  of cardinality  $|I|$ , the restriction of a code  $\mathcal{A}$  of length  $n$  is the subset of  $GF(q)^{|I|}$  defined as:

$$\mathcal{A}_I \stackrel{\text{def}}{=} \left\{ \mathbf{v} \in GF(q)^{|I|} \mid \exists \mathbf{a} \in \mathcal{A}, \mathbf{v} = (a_i)_{i \in I} \right\}.$$

The results about the unusual dimension of the square of a Reed-Solomon codes which are given in Section 2 prompt us to study the dimension of the square code  $\langle \mathcal{C}^2 \rangle$  or more generally the dimension of  $\langle \mathcal{C}_I^2 \rangle$ . When  $I$  contains no positions in  $L$ , then  $\mathcal{C}_I$  is nothing but a generalized Reed-Solomon code and we expect a dimension of  $2k - 1$  when  $|I|$  is larger than  $2k - 1$ . On the other hand, when there are positions in  $I$  which also belong to  $L$  we expect the dimension to become bigger and the dimension of  $\langle \mathcal{C}^2 \rangle$  to behave as an increasing function of  $|I \cap L|$ . This is exactly what happens as shown in the proposition below.

**Proposition 4.** Let  $I$  be a subset of  $\{1, \dots, n\}$  and set  $J \stackrel{\text{def}}{=} I \cap L$ . If the cardinality of  $I$  and  $J$  satisfy  $|J| \leq \ell - 1$  and  $|I| - |J| \geq 2k$  then

$$\dim(\langle \mathcal{C}_I^2 \rangle) = 2k - 1 + |J|. \quad (3)$$

The proof of this proposition can be found in Appendix A. An attacker can exploit this proposition to mount a distinguisher that recognizes whether a given position belongs to the secret set  $L$ . At first a set  $I$  which satisfies with high probability the assumptions of Proposition 4 is randomly chosen. Take for instance  $|I| = 3k$ . Then  $d_I \stackrel{\text{def}}{=} \dim(\langle \mathcal{C}_I^2 \rangle)$  is computed. Next, one element  $x$  is removed from  $I$  to get a new set  $I'$  and  $d_{I'} = \dim(\langle \mathcal{C}_{I'}^2 \rangle)$  is computed. The only two possible cases are then:

1. if  $x \notin L$  then  $d_{I'} = d_I$
2. and if  $x \in L$  then  $d_{I'} = d_I - 1$ .

By repeating this procedure, the whole set  $J = I \cap L$  is easily recovered. The next step now is to find all the elements of  $L$  that are not in  $I$ . One solution is to exchange one element in  $I \setminus J$  by another element in  $\{1, \dots, n\} \setminus I$  and compare the values of  $d_I$ . If it increases, it means that the new element belongs to  $L$ . At the end of this procedure the set  $L$  is totally recovered. This probabilistic algorithm is obviously of polynomial time complexity and breaks completely the homomorphic scheme suggested in [BL12].

## References

- [AAPS11] F. Armknecht, D. Augot, L. Perret, and A.R. Sadeghi. On constructing homomorphic encryption schemes from coding theory. In *Cryptography and Coding - 13th IMA International Conference, IMACC 2011, Oxford, UK, December 12-15, 2011. Proceedings*, volume 7089 of *Lecture Notes in Computer Science*, pages 23–40. Springer, 2011.
- [BL12] A. Bogdanov and C.H. Lee. Homomorphic encryption from codes, 2012.
- [CFS01] N. T. Courtois, M. Finiasz, and N. Sendrier. How to achieve a McEliece-based digital signature scheme. *Lecture Notes in Computer Science*, 2248:157–174, 2001.
- [Dal07] Léonard Dallot. Towards a concrete security proof of Courtois, Finiasz and Sendrier signature scheme. In *WEWoRC*, pages 65–77, 2007.
- [DMQN09] Rafael Dowsley, Jörn Müller-Quade, and Anderson C. A. Nascimento. A CCA2 secure public key encryption scheme based on the McEliece assumptions in the standard model. In *CT-RSA*, pages 240–251, 2009.
- [FGO<sup>+</sup>11] J.-C. Faugère, V. Gauthier, A. Otmani, L. Perret, and J.-P. Tillich. A distinguisher for high rate McEliece cryptosystems. In *Proceedings of the Information Theory Workshop 2011, ITW 2011*, pages 282–286, Paraty, Brasil, 2011.

- [Gen09] Craig Gentry. Fully homomorphic encryption using ideal lattices. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 169–178. ACM, 2009.
- [Gib91] J. Gibson. Equivalent goppa codes and trapdoors to mceliece public key cryptosystem. In Donald Davies, editor, *Advances in Cryptology EUROCRYPT 91*, volume 547 of *Lecture Notes in Computer Science*, pages 517–521. Springer Berlin / Heidelberg, 1991.
- [Köt92] R. Kötter. A unified description of an error locating procedure for linear codes. In *Proc. Algebraic and Combinatorial Coding Theory*, pages 113–117, Voneshta Voda, 1992.
- [LS01] P. Loidreau and N. Sendrier. Weak keys in the mceliece public-key cryptosystem. *IEEE Transactions on Information Theory*, 47(3):1207–1211, 2001.
- [McE78] R. J. McEliece. *A Public-Key System Based on Algebraic Coding Theory*, pages 114–116. Jet Propulsion Lab, 1978. DSN Progress Report 44.
- [MCP12] Irene Márquez-Corbella and Ruud Pellikaan. Error-correcting pairs for a public-key cryptosystem. preprint, 2012. preprint.
- [MS86] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, Amsterdam, fifth edition, 1986.
- [Nie86] H. Niederreiter. Knapsack-type cryptosystems and algebraic coding theory. *Problems Control Inform. Theory*, 15(2):159–166, 1986.
- [NIKM08] Ryo Nojima, Hideki Imai, Kazukuni Kobara, and Kirill Morozov. Semantic security for the McEliece cryptosystem without random oracles. *Des. Codes Cryptography*, 49(1-3):289–305, 2008.
- [Pel92] Ruud Pellikaan. On decoding by error location and dependent sets of error positions. *Discrete Mathematics*, 106-107:368–381, 1992.
- [RAD78] R. Rivest, L. Adleman, and M. Dertouzos. On data banks and privacy homomorphisms. *Foundations of Secure Computation*, pages 169–177, 1978.
- [SS92] V.M. Sidelnikov and S.O. Shestakov. On the insecurity of cryptosystems based on generalized Reed-Solomon codes. *Discrete Mathematics and Applications*, 1(4):439–444, 1992.
- [Wie10] C. Wieschebrink. Cryptanalysis of the niederreiter public key scheme based on grs subcodes. In N. Sendrier, editor, *Post-Quantum Cryptography, Third International Workshop, PQCrypto 2010*, volume 6061 of *Lecture Notes in Computer Science*, pages 61–72, Darmstadt, Germany, May 2010. Springer.



## A Proof of Proposition 4

The proof of Proposition 4 proceeds by exhibiting a basis of the linear space  $\langle \mathcal{C}_I^2 \rangle$ . For this purpose we define for any  $t$  in  $\{1, \dots, k\}$ ,  $\mathbf{X}^t \stackrel{\text{def}}{=} (x_i^t)_{i \in I}$  and  $\mathbf{Y}^t \stackrel{\text{def}}{=} (Y_i^t)_{i \in I}$  with:

$$Y_i^t \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } i \in J \\ x_i^t & \text{if } i \in I \setminus J. \end{cases}$$

Notice that  $\mathcal{C}_I$  is the vector space spanned by the  $\mathbf{X}^t$ 's for  $1 \leq t \leq \ell$  and the  $\mathbf{Y}^t$ 's for  $\ell + 1 \leq t \leq k$ .

The proof of Proposition 4 starts by giving a generating set for  $\langle \mathcal{C}_I^2 \rangle$ .

**Lemma 1.**  $\langle \mathcal{C}_I^2 \rangle$  is generated by the set of vectors  $\mathbf{X}^t$  for  $2 \leq t \leq 2\ell$  and  $\mathbf{Y}^t$  for  $\ell + 2 \leq t \leq 2k$ .

*Proof.* Let us define:

$$\mathbf{Z}^t \stackrel{\text{def}}{=} \begin{cases} \mathbf{X}^t & \text{if } 1 \leq t \leq \ell \\ \mathbf{Y}^t & \text{if } \ell + 1 \leq t \leq k. \end{cases}$$

Obviously,  $\langle \mathcal{C}_I^2 \rangle$  is generated by the vectors  $\mathbf{Z}^r \star \mathbf{Z}^s$  where  $r$  and  $s$  range over  $\{1, \dots, k\}$ . We notice now that

$$\mathbf{Z}^r \star \mathbf{Z}^s \stackrel{\text{def}}{=} \begin{cases} \mathbf{X}^{r+s} & \text{if } r \text{ and } s \in \{1, \dots, \ell\} \\ \mathbf{Y}^{r+s} & \text{if } r \text{ or } s \notin \{1, \dots, \ell\}. \end{cases}$$

In particular, the following equality holds:

$$\{\mathbf{Z}^r \star \mathbf{Z}^s \mid 1 \leq r \leq k \text{ and } 1 \leq s \leq k\} = \{\mathbf{X}^t \mid 2 \leq t \leq 2\ell\} \cup \{\mathbf{Y}^t \mid \ell + 2 \leq t \leq 2k\}.$$

□

The next step is to find some linear relations between the  $\mathbf{X}^t$ 's and the  $\mathbf{Y}^t$ 's. This is achieved by

**Lemma 2.** If  $\ell + |J| + 2 \leq t \leq 2\ell$ , then  $\mathbf{X}^t$  belongs to the vector space generated by

$$\bigcup_{u=\ell+2}^{t-1} \{\mathbf{X}^u, \mathbf{Y}^u\} \cup \{\mathbf{Y}^t\}.$$

*Proof.* We consider  $U$  as an indeterminate and we define the polynomials  $\varphi(U)$  and  $R(U)$  as  $\varphi(U) \stackrel{\text{def}}{=} \prod_{i \in J} (U - x_i)$  and  $R(U) \stackrel{\text{def}}{=} \varphi(U)U^{t-|J|}$ . The degree of  $R(U)$  is equal to  $t$  and hence satisfies  $\deg(R) \leq$

2 $\ell$ .  $R(U)$  can also be viewed as the polynomial  $\sum_{s=t-|J|}^t r_s U^s$  where each  $r_s$  belongs to  $GF(q)$  and

$r_t = 1$ . One can see that by construction of  $R(U)$  when  $i \in J$  then  $R(x_i) = \sum_{s=t-|J|}^t r_s x_i^s = 0$ . So if

we denote by  $\mathbf{X}_i^s$  (resp.  $\mathbf{Y}_i^s$ ) the entry of  $\mathbf{X}^s$  (resp.  $\mathbf{Y}^s$ ) at position  $i$  we equivalently have when  $i \in J$ :

$$\sum_{s=t-|J|}^t r_s \mathbf{X}_i^s = \sum_{s=t-|J|}^t r_s x_i^s = R(x_i) = 0.$$

By the very definition of  $\mathbf{Y}_i^s$  which is equal to 0 when  $i \in J$ , we have that  $\sum_{s=t-|J|}^t r_s \mathbf{Y}_i^s = 0$ . On the other hand by definition of  $\mathbf{X}^s$  and  $\mathbf{Y}^s$ , we also have that

$$\sum_{s=t-|J|}^t r_s \mathbf{X}_i^s = \sum_{s=t-|J|}^t r_s \mathbf{Y}_i^s$$

for  $i$  in  $I \setminus J$ . Therefore in all cases we have

$$\sum_{s=t-|J|}^t r_s \mathbf{X}_i^s = \sum_{s=t-|J|}^t r_s \mathbf{Y}_i^s,$$

and since  $r_t = 1$  we can write that:

$$\mathbf{X}^t = \sum_{s=t-|J|}^t r_s \mathbf{Y}^s - \sum_{s=t-|J|}^{t-1} r_s \mathbf{X}^s.$$

This concludes the proof of the lemma by noticing that  $t \geq |J| + \ell + 2$  implies that the  $s$  which appears in the sum above is larger than or equal to  $\ell + 2$ .  $\square$

It remains to prove that the generating set obtained by removing the linear relations obtained in Lemma 2 is now an independent set.

**Proposition 5.** *Assume that  $|J| \leq \ell - 1$  and  $|I| - |J| \geq 2k$ , then the set of  $\mathbf{X}^t$ 's with  $2 \leq t \leq \ell + |J| + 1$  and  $\mathbf{Y}^t$ 's with  $\ell + 2 \leq t \leq 2k$  form a basis of  $\langle \mathcal{C}_I^2 \rangle$ .*

*Proof.* A consequence of Lemma 2 is that  $\mathbf{X}^t$  with  $2 \leq t \leq \ell + |J| + 1$  and  $\mathbf{Y}^t$  with  $\ell + 2 \leq t \leq 2k$  generate the code  $\langle \mathcal{C}_I^2 \rangle$  but it remains to prove that they are linearly independent. For this purpose, let us assume that there exists a linear relation between them *i.e.*, there exist  $a_s$  and  $b_s$  in  $GF(q)$  for  $2 \leq s \leq 2k$  such that:

$$\sum_{s=2}^{\ell+|J|+1} a_s \mathbf{X}^s + \sum_{s=\ell+2}^{2k} b_s \mathbf{Y}^s = 0. \quad (4)$$

By setting  $a_s = 0$  for  $\ell + |J| + 2 \leq s \leq 2k$  and  $b_s = 0$  for  $2 \leq s \leq \ell + 1$ , Equation (4) can be rewritten as:

$$\sum_{s=2}^{2k} (a_s \mathbf{X}^s + b_s \mathbf{Y}^s) = 0. \quad (5)$$

Let us denote  $R(U) \stackrel{\text{def}}{=} \sum_{s=2}^{2k} (a_s + b_s) U^s$ . We know that if  $i \notin J$  then  $\mathbf{Y}_i^s = \mathbf{X}_i^s = x_i^s$  for  $s$  in  $\{2, \dots, 2k\}$ . Therefore by Equation (5) we have  $R(x_i) = 0$  for any  $i \notin J$ . As we have assumed that  $|I| - |J| \geq 2k$ , it implies that  $R(U) = 0$  or equivalently  $a_s = -b_s$  for all  $s$ . In particular  $a_s = 0$  for  $2 \leq s \leq \ell + 1$ . On the other hand, when  $i \in J$ , we have  $\mathbf{Y}_i^s = 0$  and  $\mathbf{X}_i^s = x_i^s$  for any  $s$  in  $\{\ell + 2, \dots, \ell + |J| + 1\}$ . Hence when  $i \in J$ , Equation (4) leads in fact to:

$$\sum_{s=\ell+2}^{\ell+|J|+1} a_s \mathbf{X}_i^s = 0. \quad (6)$$

Now let us consider  $Q(U) \stackrel{\text{def}}{=} \sum_{s=\ell+2}^{\ell+|J|+1} a_s U^s$  and observe that there exists some polynomial  $S(U)$  with  $\deg(S) \leq |J| - 1$  such that:

$$Q(U) = U^{\ell+2} S(U).$$

From Equation (6) we know that  $Q(x_i) = 0$  for all  $i$  in  $J$ . Since all  $x_i$ 's are different from 0, this implies that  $S(x_i) = 0$ . Since  $\deg(S) \leq |J| - 1$  this means that  $S(U) = 0$ , and therefore  $a_s = 0$  for all  $\ell + 2 \leq s \leq \ell + |J| + 1$ . Then equation (4) holds if and only if all the coefficients  $a_s$  and  $b_s$  are zero, which means that  $\mathbf{X}^t$  with  $2 \leq t \leq \ell + |J| + 1$  and  $\mathbf{Y}^t$  with  $\ell + 2 \leq t \leq 2k$  form indeed a basis of  $\langle \mathcal{C}_I^2 \rangle$  whose dimension is therefore  $2k - 1 + |J|$ .  $\square$

Proposition 4 immediately follows from Proposition 5 which characterises a basis of  $\langle \mathcal{C}_I^2 \rangle$ .